SKIN INFECTION DETECTION SYSTEM USING ML

Bhushan Chaudhari1, Rahul Jeevan Patil2, Prathamesh Shinde3, Rohan Suryawanshi4, Ashish Lulla5

*1 (Associate professor, Dept of Information Technology SVKM’s IOT , Dhule India,* [*chaudharibs@gmail.com*](mailto:chaudharibs@gmail.com)*\*)*

*2345 (BTECH IV year, Dept of Information Technology SVKM’s IOT , Dhule India,rahuljjevanpatil@gmail.com)*

**Abstract**

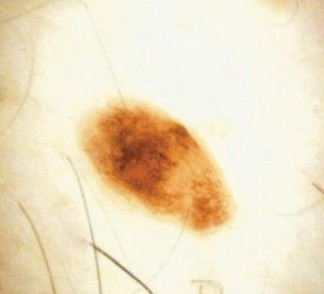
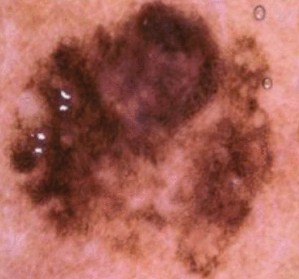
**Skin cancer is a hazardous type of cancer that can be fatal. To improve the prognosis, skin cancer must be diagnosed early in its early stages. Early skin cancer detection is difficult and costly, though. Cancerous tumors can be difficult to diagnose even when they are apparent since they might resemble benign lesions quite a bit. Examining all pigmented skin lesions surgically can be uncomfortable and cause scars. Thus, the requirement for an automatic and**

**and painless skin cancer detection system with high accuracy. Machine Learning (ML) and Deep Learning (DL) have shown promise in skin cancer detection. This paper compares the effectiveness of several DL models, including ResNetv2, VGG16, EfficientNet-B5, and EfficientNet-B7, using pre-trained Convolutional Neural Networks (CNNs), with a ML model, Support Vector Machine (SVM), to determine if a skin sample is cancerous. The results demonstrate that the CNN models outperform the SVM in accuracy, precision, recall, and F1-score, with EfficientNet-B7 achieving the highest F1-score of 84.22%...**

*Keywords —skin cancer, SVM, CNN, Deep Learning, Machine learning.*

1. **Introduction**

It has been acknowledged that skin cancer is one of the most common illnesses in the globe since the 1970s. It is defined by an extremely great multiplication of skin cells that are biologically abnormal. Melanoma is the most common kind of skin cancer that can be triggered by malignant tumors, which have a tendency to spread rapidly. Over the past few decades, there was an obvious increase in the number of persons diagnosed with both melanoma and non-melanoma skin cancer. A significant portion of skin cancer situations are triggered by melanoma, in particular, which is a serious condition. The World Health Organization (WHO) predicts that one in three Americans will be affected by skin cancer in the next years. A separate research carried out by the Skin Cancer Foundation (SCF) indicates that this percentage might rise to one in five persons. Skin cancer incidence has also increased in western nations as well, such as Australia and Canada. Skin cancer, especially melanoma and non-melanoma variations, is a severe health concern that may severely hamper normal physical processes. In 2018, there were one million incidences of non-melanoma skin cancer and over 300,000 cases of melanoma globally. It is expected that these figures will rise to 2,450,000 in the next 20 years. Mortality from skin cancer is closely linked to the disease's stage. For instance, the overall survival rate for people with melanoma is around 98.4%; however, it drops to 22.5% in cases of metastatic melanoma. Early identification and treatment is critical because as the condition worsens, the death rate tends to rise. Dermatologists usually make the diagnosis of skin cancer by examining the patient's skin for benign or malignant lesions. But it can be difficult to differentiate between the two kinds of lesions; therefore, a radiologist's skill is required for a precise diagnosis. Numerous reasons, such as viruses, inflammation, environmental factors, allergies, and infections, can result in skin cancer. Although there are therapies for skin cancer, those who receive a diagnosis early on tend to have the highest survival rates. The skin acts as a barrier to protect the body's internal organs, and any abnormalities in skin function can point to underlying problems, including skin cancer. Early discovery of cancer of the skin stages can enhance the outcome of patients, hence it is so important to precisely diagnose the stage of cancer. A delayed diagnosis can drastically diminish the likelihood of survival. Numerous techniques, including support vector machines (SVM), Neural Networks (NN), K-Nearest Neighbors (KNN), and Naive Bayes (NB), have been employed by scientists to forecast the disease's early stages. Low error rates, high efficiency, and accurate prediction outcomes have been achieved with these methods. Machine learning (ML) and deep learning (DL) models are used in many different areas, such as robotics, healthcare, hospitality, and agriculture, due to their better prediction of outcomes skills. For the purpose of to predict skin cancer, this study used a variety of pre-trained algorithms, especially those that classified pictures either as harmless or cancerous. Four different classifiers were applied in order to accomplish the classification: SVM, KNN, NB, and NN.

Benign Malignant

**Figure 1 Sample image of Benign and Malignant lesion**

1. **Literature Survey**

To identify skin cancer, researchers used KNN, SVM, and NB classifiers [13]. The International Skin Imaging Collaboration provided the 672 pictures of skin lesions which made up their dataset (ISIC). The corresponding sensitivity and specificity rates for SVM, KNN, and NB were 75%, 65%, 85%, 86.2%, and 82%, 72% for each model.

In another investigation [14], scientists used CNN, AI, and SVM to build a hybrid model for melanoma detection. This model beat other existing models, with a total precision of 85%.

In addition, a CNN DL model was used to detect skin cancer [6]. The study's 2018 dataset, which included 10,015 images of seven different skin lesion illnesses, was made possible via the ISIC alliance. Six data augmentation techniques were applied to boost the data set's size.

In [15], a DL approach was used to detect skin lesions at an early stage. Region-based CNN and fuzzy-based K-means algorithms were applied to datasets with some noise; the following accuracy rates were obtained for the ISIC-2016, PH2, and ISIC-2017 datasets: 95.40%, 95.6%, and 93.1%, respectively.

In a separate research [16], skin cancer was identified from a dataset of 21,804 photographs of benign and malignant carcinoma using a CNN-based VGG-16 model. The authors experimented with other picture sizes, including 299299 and 224224 pixels, and they found only little variations in the results.

In [4], three types of skin cancer—bearing, benign, and malignant—were investigated using CNN and conventional ML models. The results of the investigation showed that the CNN model performed better than more traditional machine learning technique

SVM and random forest algorithms are used in the hybrid skin cancer detection model developed by the authors [17]. Features were extracted from the pictures after they had been segmented and utilized for classification. The hybrid model performs better than each model alone, as seen by its accuracy in the Grey level co-occurrence matrix (88.56%) compared to 85.72% for SVM.

In [5], utilizing the HAM 10000 dataset, which included 10,015 cancer images, a DL model named DCNN was created for the purpose of categorizing malignant and benign cancer. Comparing the DCNN model's performance to that of the AlexNet, VGG16, ResNet, MobileNet, and DenseNet models showed how resilient it was.

The authors' hybrid skin cancer detection model combines SVM and random forest techniques [17]. After the photos were split, features were taken out and used for categorization. As demonstrated by its accuracy in the Grey level co-occurrence matrix (88.56%) as opposed to SVM's 85.72%, the hybrid model outperforms each model alone.

.

A DL model named DCNN was created in [5] to use the HAM 10000 dataset, which had 10,015 cancer images, to identify malignant and benign cancer. By contrasting its performance with that of the AlexNet, VGG16, ResNet, MobileNet, and DenseNet models, the robustness of the DCNN model was illustrated.

Another study [3] used three algorithms: Bayesian ensembling, Monte Carlo dropout, and Spectral Normalized Neural Gaussian Process for skin cancer prediction. Using a pre-trained VGG16 model, the accuracy rates for the Bayesian ensembling approach varied from 66% to 69%.

Lastly, a CNN classifier was employed in [1] to classify nodular melanoma, superficial spreading, and lesion melanoma using a dataset from the dermnetnz website. The CNN model outperformed the decision tree, random forest, and gradient boosting models, with 91.07% accuracy compared to 70%, 71%, and 68% for the latter three.

* 1. **Methodology**

In machine learning there are two different modes namely training and testing mode.

* 1. **dataset**

The investigation employed a dataset on skin cancer that was gathered from [19]. 10,605 photos showing the benign and malignant stages of skin cancer are included in this dataset. The dataset is divided into two files, named "train" and "test," with 9,605 photos in the former folder (5,000 benign and 4,605 malignant), and 500 benign and 500 malignant images in the latter.

* 1. **TRAINING AND TESTING MODELS** **Training mode**

|  |  |
| --- | --- |
| Manually obtain the characteristics of skin cancer | |
|  |  |
| Calculate area of spread | |
|  |  |
| Based on area ,label the stages of images | |

Read the image

Repeat the process for some images

# Fig. 2 Training Mode

**Testing Mode**

|  |  |
| --- | --- |
| Read the image | |
|  |  |
| automatically extract the characteristics of image | |
|  |  |
| Apply characteristics to machine learning | |

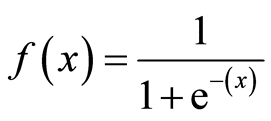
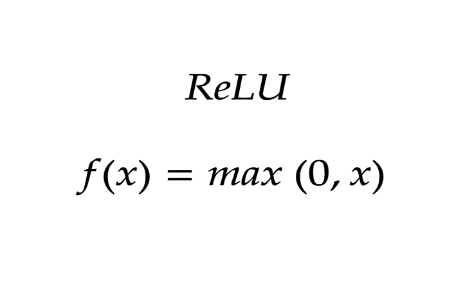
Display stage based on the outcome of the eq

# Fig. 2.1Testing Mode

The training step is essential after the model is set up, and in this case, it was done using Keras and TensorFlow across 50 epochs. Using a collection of infrared images from cyclone satellites, the system was trained to projected wind speed by minimizing the loss function (MSE).

The probability of tumor growth that stays confined and does not spread to other regions of the body is indicated by the benign stage of skin cancer. Nevertheless, skin cancer that grows to a malignant stage has the risk of spreading and affecting other organs, including the brain, liver, lungs, and bones. Early detection and management of malignant cancer are essential to stopping its spread. Chemotherapy, which involves putting chemicals into the veins to kill cancer cells, is one type of cancer treatment. However, the effectiveness of this treatment depends on quick detection of skin cancer. Among the unpleasant side effects of chemotherapy include fatigue, hair loss, nausea, and diarrhea. Applying ML and DL models is therefore essential for rapid skin cancer diagnosis.

A separate collection of hyperparameters were chosen for every method. For example, ReLu, Adam, 0.0001, 200, and 100 were the values for the activation function, learning rate, number of iterations, and number of hidden layers of the NN model. The kernel chosen was RBF, and the SVM iteration count was set to 100. The values provided to the metric, weight, and amount of neighbors in KNN were five Euclidean, uniform, and number of neighbors. Finally, N does not require any particular hyper-parameters to be created.

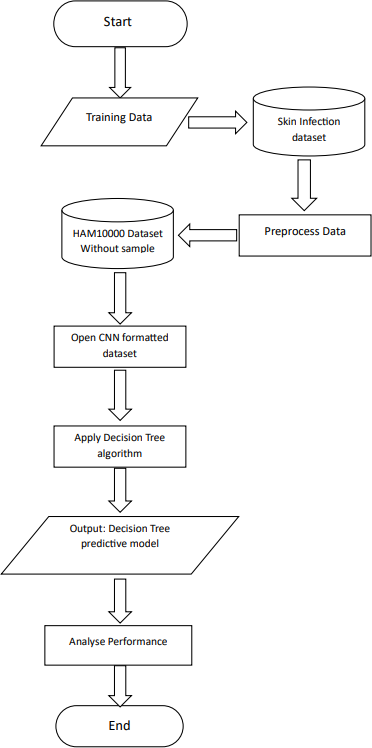


# Activation function equation: ReLu

The study included four models for the classification of skin cancer: Naive Bayes, K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Neural Network (NN), based on results from a comprehensive research survey. The models in question were chosen as they are regarded as some of the best options available for classification-related difficulties.

Originally, 90% of the dataset was set aside for training, while the other 10% was reserved for testing. The pre-trained SqueezeNet, an 18-layered CNN model renowned for its effectiveness in identifying and categorizing features applications, was used for feature extraction. It was able to categorize items into up to 1000 classes using Squeeze Net’s 2016 form.

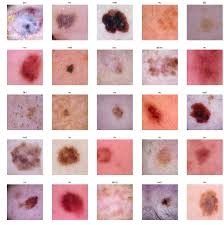
**UML Diagram:**



# Fig.3 UML diagram

1. **Result and Discussion**

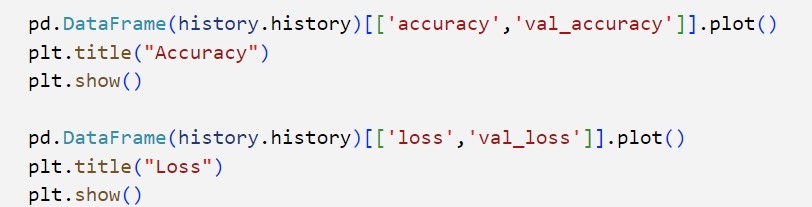
**Details of Dataset:**

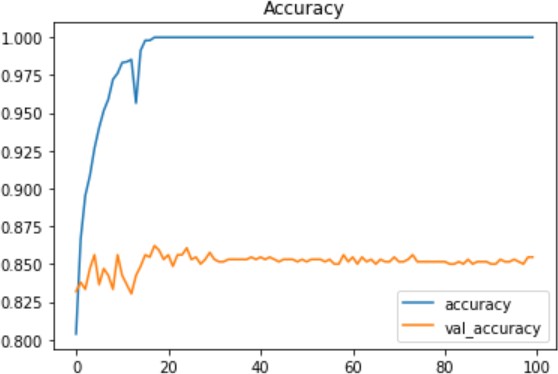


# Fig.4 Data Set

**Experimentation and result**

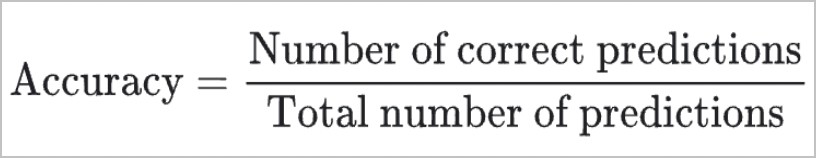




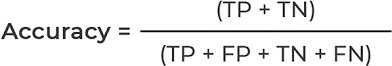


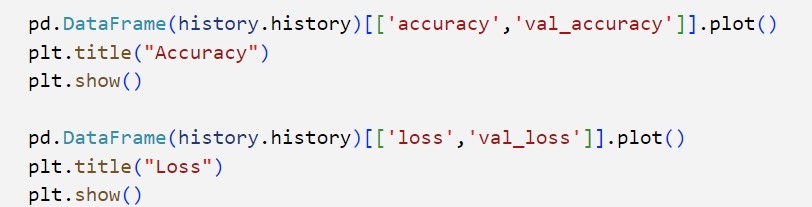
# Fig. 5 Model Accuracy

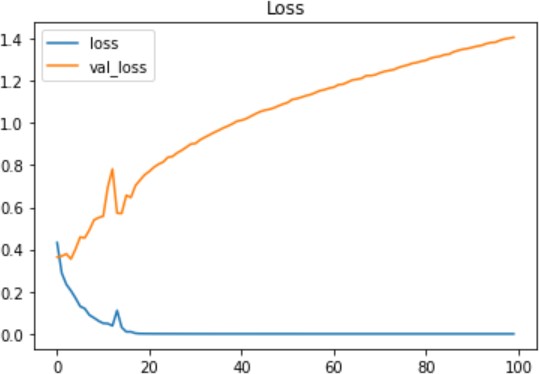
The pyplot API in matplotlib, which was first developed as a free successor to MATLAB, provides an easy stateful interface akin to MATLAB. Sometimes individuals consider the Object-Oriented (OO) API to be more complicated to use, even if it is more powerful and adaptable. The pyplot interface thus becomes frequently utilized and used as the default.



In terms of both positive and negative outcomes, accuracy for binary categorization can also be calculated using the following method:



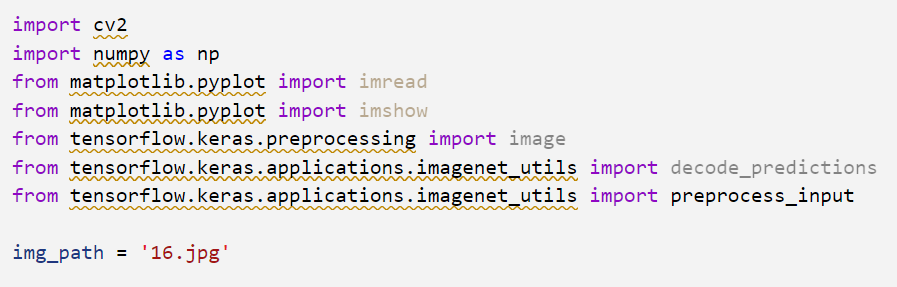
Where, TP = TRUE POSITIVES, TN = TRUE NEGATIVES , FP = FALSE POSITIVES AND FN = FALSE NEGATIVES.



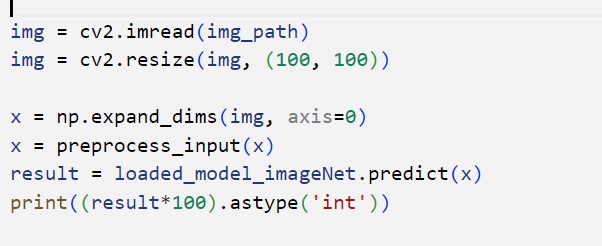
# Fig.6 Model Loss

Machines can acquire novel skills through lowering a loss function that assesses how well an algorithm illustrates the given data. When predictions and actual results diverge considerably, a high number is generated by the loss function. In time, optimization processes enable the loss function to reduce errors in prediction. This article covers the various loss coefficients and their uses in machine/deep learning.

Almost each algorithm in machine learning utilizes the exact same function of loss. The choice of loss function is affected by the kind of machine learning method, how simple it is to calculate derivatives, and, to a certain extent, whether or not there are any outliers in the dataset



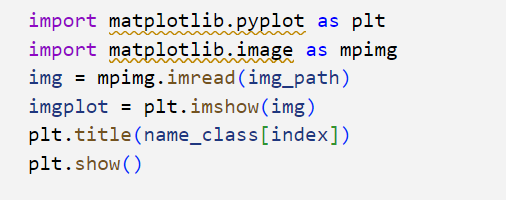
# loading an Image



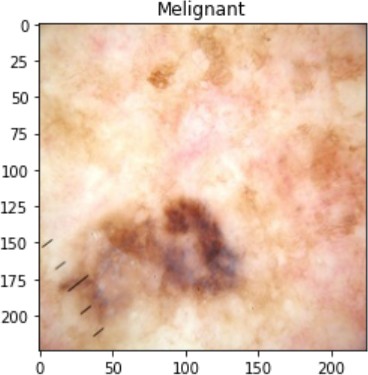
**Prediction**



# Calculation



**Plotting**



# Fig. 7Actual Result

**Conclusion:**

The need for a dependable and efficient automated method for diagnosing skin cancer is growing increasingly obvious in view of the increasing incidence of skin cancer in recent decades. This work uses the MobileNet model, which was trained on 38,569 dermoscopy as pictures from the HAM10000 dataset, to demonstrate how effective deep learning is in automatic dermoscopic multi-class skin cancer classification. For seven categories, the model's overall precision was 83.1%; its top two and top three accuracies were 91.36% and 95.34%, respectively. The accuracy, recall, and F1-score weighted averages were 89%, 83%, and 83%, respectively, which equaled the results of dermatologists with a thorough education on seven diagnostic tests.

The study leads to the opinion that an effective real-time computer-aided system for automated medical diagnosis may be created with the MobileNet idea. Along with its quicker and smaller design, MobileNet showed precise and robust performance when contrasted with previous models. In order to enhance the accuracy and efficacy of computer-aided systems for the diagnosis of skin cancer, future research may entail the incorporation of individual patient data, including genes, age, and skin color.

**References**

1. The International Journal of Computing, Programming and Database Management, vol. 3, no. 1, pp. 141–147, 2022; V. R. Allugunti, "A machine learning model for skin disease classification using convolution neural network."
2. At the WSCG'2021 - 29. International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2021, 2021, M. S. Junayed, N. Anjum, A. Sakib, and M. B. Islam presented "A deep CNN model for skin cancer detection and classification." doi: 10.24132/ csrn.2021.3101.8.
3. "Uncertainty-aware skin cancer detection: The element of doubt," P. Tabarisaadi, A. Khosravi, and S. Nahavandi, Comput. Biol. Med., vol. 144, p. 105357, May 2022.

[4]T. Saba, "Handcrafted and non-handcrafted features in computer vision for microscopic skin cancer diagnosis," Microsc. Res. Tech., vol. 84, no. 6, pp. 1272–1283, Jun. 2021.

1. "An enhanced technique of skin cancer classification using deep convolutional neural network with transfer learning models," M. S. Ali, M. S. Miah, J. Haque, M. M. Rahman, and M. K. Islam, Machine Learning with Applications, vol. 5, p. 100036, Sep. 2021.
2. “Deep Learning Solutions for Skin Cancer Detection and Diagnosis," by H. Nahata and S. P. Singh, in

V. Jain and J. M. Chatterjee, eds., Machine Learning and Healthcare: Machine Learning and Healthcare, Cham: Springer International Publishing, 2020, pp. 159–182.

1. “A deep learning based convolutional neural network model with VGG16 feature extractor for the detection of Alzheimer Disease using MRI scans," S. Sharma, K. Guleria, S. Tiwari, and S. Kumar, Measurement: Sensors, vol. 24, p. 100506, Dec. 2022.
2. K. Guleria, S. Sharma, S. Kumar, and S. Tiwari, "Predictive machine learning and deep learning for early hypothyroidism and multiclass classification," Measurement: Sensors, vol. 24, p. 100482, Dec. 2022.
3. In 2022 International Conference on Decision Aid Sciences and Applications (DASA), March 1067– 1072, S. Sharma, A. Kataria, and J. K. Sandhu, "Applications, Tools and Technologies of Robotic Process Automation in Various Industries," pp. 1067–1072.
4. “Machine Learning Techniques and Implementation of Different ML Algorithms," S. Singh, K. R. Ramkumar, and A. Kukkar, in 2021 2nd Global Conference for Advancement in Technology (GCAT), Oct. 2021, pp. 1–6.
5. "A framework for hotel inventory control system for online travel agency," by A. Sharma and K.

Guleria

1. "Rice Leaf Blight Disease detection using multi-classification deep learning model," 10th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 2022, pp. 1–5, doi: 10.1109/ICRITO56286.2022.9964644, R. Sharma, V. Kukreja,

R. K. Kaushal, A. Bansal, and A. Kaur.

1. In the 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), July 2020, pp. 1–5, M. Vidya and M. V. Karki, "Skin Cancer Detection using Machine Learning Techniques."
2. M. M. Vijayalakshmi, "Image processing and machine learning approaches for the detection of melanoma skin cancer," International Journal of Trend in Scientific Research and Development (IJTSRD), vol. 3, no. 4, pp. 780–784, 2019.
3. M. Nawaz et al., "Deep learning and fuzzy k-means clustering for skin cancer detection from dermoscopic images," Microsc. Res. Tech., vol. 85, no. 1, pp. 339–351, Jan. 2022.
4. Diagnostics (Basel), vol. 12, no. 1, Dec. 2021, doi: 10.3390/diagnostics12010040; M. Nauta, R. Walsh,

A. Dubowski, and C. Seifert, "Uncovering and Correcting Shortcut Learning in Machine Learning Models for Skin Cancer Diagnosis."

1. "Diagnosis of skin cancer using machine learning techniques," Micro-process. Microsyst., vol. 81, p. 103727, Mar. 2021, A. Murugan, S. A. H. Nair, A. A. P. Preethi, and K. P. S. Kumar.
2. The paper "Skin Cancer Classification Using Image Processing and Machine Learning" was presented by A. Javaid, M. Sadiq, and F. Akram at the 2021 International Bhurban Conference on Applied Sciences and Technologies (IBCAST) in January 2021, pages 439–444.
3. "Melanoma Skin Cancer Dataset of 10000 images," by M. H. Javid 29 March 2022. Reached on November 18, 2022. [Online]. Accessible:
4. A. Keutzer, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and F. N. Iandola, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and...